Basu+Ghosh, 有限母集団からのサンプリングにおける十分統計量, の紹介 (Invariant Statistical ProblemsとSufficiency)

渋谷, 政昭

数理解析研究所講究録
46巻 1968年 47-54

Basu + Ghosh，有限母集団からのサンプリングにおける十分統計量，の紹介

統数研　渋谷政昭

§1. 確率分布族が undominated の場合についての十分統計量の概念を論じるためには十分統計量の定義を拡張しなければならない．しかしながら，応用統計学で扱う確率分布族はかなり限定されているので，簡単な枠内で扱えることを示すのが，この論文

D. Basu & J. K. Ghosh, Sufficient statistics in sampling from a finite universe, 36 th sess. Intnl. Statist. Inst.., 1967, Sydney

の目的である．論文の評価については別稿森本治樹の remark を見よ．

典型的な実例としていわゆる有限母集団からのサンプリングを考える．N個の対象の特性を

$$\Theta = \{ \theta_1, \cdots, \theta_N \}, \quad -\infty < \theta_j < \infty$$

とする．Θ（の関数）についての推測を行うのに，ある大きさの標本をある確率方式にしたがって選ぶ．たとえば1個を

等確率に選ぶとすると，標本空間は E' で，確率分布は N個の点 θ の上の等確率，$1/N$，分布である．これは σ 有限の測度について dominated とならない．θ を標本空間とみなすことは標本空間が'パラメータ θ'に依存することになり，不適切な定式化である．

§2．$(\mathcal{X}, \mathcal{A}, \mathcal{P})$ を，標本空間，σ-field，確率測度族とする．'統計量'を標本空間のある分割；

$$\Pi = \{ \pi_t ,\ t \in T \} \qquad \bigcup_{t \in T} \pi_t = \mathcal{X}$$

と定義する．一般にTは可算でなく，π_t は $\mathcal{A}$ 可測でない．

'分割 Π が誘導する $\mathcal{A}$ の subfield'を，Π に属する部分 π_t の和集合で $\mathcal{A}$ に属するものの全体；

$$\mathcal{A}(\Pi) = \{ A ;\ A = \bigcup_{t} \pi_{t'} ,\ A \in \mathcal{A} \}$$

により定義する．

任意に与えた subfield が必ずしも分割から誘導できないことが次の例で示される：$\mathcal{A}$ をボレル集合の全体，$\mathcal{C} \subset \mathcal{A}$ をすべての可算集合とその余集合の全体とすると，$\mathcal{C}$ を誘導するような分割は各点分割，Π，でなければならないが，$\mathcal{A}(\Pi)$ は $\mathcal{A}$ そのものになってしまう．

逆に，subfield $\mathcal{B}$ が与えられたとき，任意の点 x にたいして $\pi_x = \bigcap \{ B ;\ x \in B ,\ B \in \mathcal{B} \}$ と定義し，異る π_x の

全体でできる分割を'$\mathcal{B}$から誘導される分割,' $\Pi(\mathcal{B})$, とする. subfieldから誘導されないような分割が存在することは, $\mathfrak{A}$をボレル集合の全体とし, ある $B \in \mathfrak{A}$ と B^c への分割を例とすればよい.

subfield $\mathcal{B}_1$ の要素がすべて $\mathcal{B}_2$ に含まれていれば, $\mathcal{B}_1 \subset \mathcal{B}_2$ と書く. 分割 Π_1 の各部分が Π_2 の部分の和集合になっていれば $\Pi_1 < \Pi_2$ と書く. 次の関係が容易に言える.

$$\mathcal{B}_1 \subset \mathcal{B}_2 \quad \Rightarrow \Pi(\mathcal{B}_1) < \Pi(\mathcal{B}_2)$$

$$\Pi_1 < \Pi_2 \quad \Rightarrow \mathfrak{A}(\Pi_1) \subset \mathfrak{A}(\Pi_2)$$

$$\Pi(\mathfrak{A}(\Pi)) < \Pi$$

$$\mathfrak{A}(\Pi(\mathcal{B})) \supset \mathcal{B}.$$

さて'統計量, すなわち分割Πが十分'であることを, $\mathfrak{A}(\Pi)$ が十分なことにより定義する.

§3. 上の定義がある枠内で有効であることを示す前に, 2つの病理例をあげる. いずれにおいても, $\mathcal{X} = E^1$, $\mathfrak{A}$ はボレル集合の全体, Bはある原点を含まず原点に関し対称な非ボレル集合, とする.

T. S. Pitcher ('57, Ann. Math. Statist.) の病理例:

$$\begin{cases} P_\theta(\theta) = P_\theta(-\theta) = \frac{1}{2} & , \quad \theta \in B \\ P_\theta(\theta) = 1 & , \quad \theta \bar{\in} B \end{cases}$$

する離散分布族 $\{P_\theta\}$ を考える. A を B に含まれる対称なボレル集合とすると, 統計量

$$t_A(x) = \begin{cases} |x| & x \in A \\ x & x \notin A \end{cases}$$

は十分であり, したがって最小十分統計量は存在しない.

D. C. Burkholder ('61, Ann. Math. Statist.) の病理例.

$$\begin{cases} P_\theta(\theta) = P_\theta(-\theta) = 1/2 & \theta \neq 0, \ -\infty < \theta < \infty \\ P_\theta(0) = 1 & \theta = 0 \end{cases}$$

という離散分布の族を考える. 対称なボレル集合の全体, $\mathcal{C}$, は十分な subfield である. ところが

$$\mathcal{C}^* = \{A;\ A \in \mathfrak{A},\ A \cap B \text{ が対称}\}$$

は $\mathcal{C} \subset \mathcal{C}^*$ であるにもかかわらず十分でない. なぜならば $\mathcal{C}^*$ が十分ならば, たとえば $(0, \infty)$ の条件つき確率が

$$P((0,\infty) \mid x) = \begin{cases} 1/2 & x > 0,\ x \in B \\ 1 & x > 0,\ x \notin B \end{cases}$$

とならなければならないが, これは $\mathfrak{A}$ 可測でない.

§4. 問題点はボレル可測性にあるのだが, 離散分布族を考える限りではもっと'大きい'集合族を考えてよい. われわれの問題の枠組として適当なのは次のような $(\mathfrak{X}, \mathfrak{A}, \mathcal{P})$ である:

$\mathcal{X}$：（非可算），$\mathcal{Q}$：$\mathcal{X}$の部分集合の全体

$\mathcal{P}$：离散確率測度の族（非可算）であって，

$$P(A)=0 \quad \text{for all} \quad P\in\mathcal{P} \quad \text{ならば} \quad A=\phi .$$

$P(x)$ で $P(\{x\})$ を表わすことにする．

定理1（分解定理）分割 Π が十分である必要十分条件は

$$P(x)=g(x)P(\pi_x) \qquad x\in\pi_x\in\Pi$$
$$\text{for all} \quad x\in\mathcal{X}, \quad P\in\mathcal{P}$$

と表わせることである．

証明．$\Rightarrow$ Π が十分ならば

$$(*) \qquad P(A\cap B)=\int_B f(A|\cdot)\,dP(\cdot) \quad \text{for all } A\in\mathcal{Q},\ P\in\mathcal{P}$$

なる $\mathcal{Q}(\Pi)$ 可測関数 $f(A|\cdot)$ が存在する．$A=\{x\}$，$B=\pi_x$ とする．$\mathcal{Q}$ の定義から Π の部分は $\mathcal{Q}(\Pi)$ の atom であり，したがって f が $\mathcal{Q}(\Pi)$ 可測とは各 π_x の上で定数ということになる．それを $g(x)$ と書けば定理の式となる．

$\Leftarrow$ 分解式の両辺を $x\in\pi_x$ にわたって加える．（このとき左辺 >0 となる x は可算 —— それは P によって変わるが —— ところが $g(x)>0$ となる $x\in\pi_x$ は P に依存しないから π_x そのものが可算であることになる．）すると $\sum_{x\in\pi_x} g(x)=1$．そこで，

$$f(A|x) = \sum_{y \in \pi_x \cap A} g(y)$$

と定義すると，任意の $B \in \mathfrak{A}(\Pi)$ について $(*)$ が導ける.

定理 2. もっとも粗い，十分な分割（最小十分統計量）が必ず存在する.

証明. $\mathcal{P}$ の部分分布族

$$\mathcal{P}_x = \{ P ;\ P \in \mathcal{P},\ P(x) > 0 \}$$

を定義とする. $\mathcal{P}$ についての最初の仮定から，すべての x にたいして $\mathcal{P}_x \neq \phi$.

同値関係 $x \sim y$ を

$$\mathcal{P}_x = \mathcal{P}_y \quad , \quad P(x)/P(y) = \text{const.} \quad P \in \mathcal{P}_x$$

で定義し，同値な点による分割を Π^* とする. 定理 1 より Π^* の十分性が容易にわかる.

また任意の十分な分割 Π に属する π_x の 2 点 x, y は上の意味で同値となり，したがって $\Pi^* < \Pi$.

定理 3. 十分な subfield $\mathcal{B}$ を誘導する分割が必ず存在する.

証明. $\mathfrak{A}(\Pi(\mathcal{B})) \supset \mathcal{B}$ が任意の subfield について言えるから，十分な $\mathcal{B}$ について $\subset$ を言えばよい.

1° $\mathcal{B}$ の十分性から，ある固定した $\pi \in \Pi(\mathcal{B})$ にたいし

て，$\mathcal{B}$ 可測で P によらない関数 $f(\pi|\cdot)$ が存在し，

$$P(\pi \cap B) = \int_B f(\pi|\cdot)\, dP(\cdot) \qquad \text{for all} \quad B \in \mathcal{B},\ P \in \mathcal{P}$$

$\phi(\cdot) = f(\pi|\cdot)$ と書くことにする。

$$\phi(x) = c \quad , \quad x \in \pi$$

であるが，逆に

$$C = \phi^{-1}(\{c\})$$

なる集合を考えると，$\pi \subset C \in \mathcal{B}$ である．

$$P(\pi) = P(\pi C) = \int_C \phi(\cdot)\, dP(\cdot) = c\, P(C) > 0 \quad \text{for some } P.$$

したがって $c > 0$ である。

2° 任意の $\pi_1 \in \Pi$，$\pi_1 \neq \pi$ にたいして $\phi(x) \neq c$，$x \in \pi_1$ となる：帰謬法で $\pi_1 \subset C$ とすると，π_1 を含み π を含まぬ $B \in \mathcal{B}$ が存在して

$$0 = P(\pi \cap B \cap C) = \int_{BC} \phi(\cdot)\, dP(\cdot) = c\, P(BC) \geq c\, P(\pi_1)$$

つまり，すべての P にたいして $P(\pi_1) = 0$ となり矛盾である．

結局 $\pi = C$ が言えた．

3° 任意の $B \in \mathcal{Q}(\Pi(\mathcal{B}))$ が $\mathcal{B}$ に属することをいう．

$$P(B \cap B') = \int_{B'} f(B|\cdot)\, dP(\cdot)$$

において $\pi \in \Pi$ $(\pi \in \mathcal{B})$ を B' に入れ $f(B|\cdot) = \lambda(\cdot)$ と書けば，

$$P(B \cap \pi) = \int_\pi \lambda(\cdot)\, dP(\cdot) = e\, P(\pi) = \begin{cases} P(\pi) & \pi \subset B \\ 0 & \text{otherw} \end{cases}$$

つまり πの上で一定の値をとる関数 λ が B の indicator となっている。したがって $B \in \mathcal{B}$. $\mathcal{B}$ は Π の部分の和集合全体である.

注意1. Π が十分ならば $\Pi < \Pi^*$ なる Π^* は必ず十分である.

注意2. しかしながら subfield については $\mathcal{C}$ が十分であっても $\mathcal{C} \subset \mathcal{C}^*$ なる $\mathcal{C}^*$ が十分とは言えない。

注意3. $\mathcal{B}_1$, $\mathcal{B}_2$ が十分ならば $\mathcal{B}_1 \cap \mathcal{B}_2$ が十分であることは言える.